Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2016, Vol. 52, No. 2, 888–897
DOI: 10.1214/14-AIHP658

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Strong Feller properties for degenerate SDEs with jumps

Zhao Dong[b], Xuhui Peng[a,b,1], Yulin Song[c] and Xicheng Zhang[d]

[a]*College of Mathematics and Computer Science, Hunan Normal University, Changsha 410081, P.R. China. E-mail:* pengxuhui@amss.ac.cn
[b]*Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, Beijing 100190, P.R. China. E-mail:* dzhao@amt.ac.cn
[c]*Department of Mathematics, Nanjing University, Nanjing, Jiangsu 210093, P.R. China. E-mail:* songyl@amss.ac.cn
[d]*School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, P.R. China. E-mail:* XichengZhang@gmail.com

Received 12 February 2014; revised 20 October 2014; accepted 23 October 2014

Abstract. Under full Hörmander's conditions, we prove the strong Feller property of the semigroup determined by an SDE driven by additive subordinate Brownian motions, where the drift is allowed to be arbitrary growth. For this, we extend a criterion due to Malicet and Poly (*J. Funct. Anal.* **264** (2013) 2077–2096) and Bally and Caramellino (*Electron. J. Probab.* **19** (2014) 1–33) about the convergence of the laws of Wiener functionals in total variation. Moreover, the example of a chain of coupled oscillators is verified.

Résumé. Sous des conditions de Hörmander fortes, nous prouvons la propriété forte de Feller pour le semi-groupe déterminé par une SDE dirigée par des mouvements browniens subordonnés additifs, où la dérive est autorisée à être arbitrairement croissante. Pour cela, nous étendons un critère dû à Malicet et Poly (*J. Funct. Anal.* **264** (2013) 2077–2096) et à Bally et Caramellino (*Electron. J. Probab.* **19** (2014) 1–33) sur la convergence, en variation totale, des lois de fonctionnelles de Wiener. Ce résultat couvre le cas d'une chaîne d'oscillateurs couplés.

Keywords: Strong Feller property; SDE; Malliavin's calculus; Cylindrical α-stable process; Hörmander's condition

1. Introduction

Let $\mathbb{W}$ be the space of all continuous functions from $\mathbb{R}_+ := [0,\infty)$ to $\mathbb{R}^m$ vanishing at the starting point 0, which is endowed with the locally uniform convergence topology and the Wiener measure $\mu_{\mathbb{W}}$ so that the coordinate process $W_t(\omega) = \omega_t$ is a standard m-dimensional Brownian motion. Let $\mathbb{H} \subset \mathbb{W}$ be the Cameron–Martin space consisting of all absolutely continuous functions with square integrable derivatives. The inner product in $\mathbb{H}$ is denoted by

$$\langle h_1, h_2\rangle_{\mathbb{H}} := \sum_{i=1}^m \int_0^\infty \dot{h}_1^i(s)\dot{h}_2^i(s)\,\mathrm{d}s.$$

The triple $(\mathbb{W}, \mathbb{H}, \mu_{\mathbb{W}})$ is also called the classical Wiener space.

Let D be the Malliavin derivative operator. For $k \in \mathbb{N}$ and $p \geq 1$, let $\mathbb{D}^{k,p}$ be the associated Wiener–Sobolev space with the norm:

$$\|F\|_{k,p} := \|F\|_p + \|DF\|_p + \cdots + \left\|D^k F\right\|_p,$$

where $\|\cdot\|_p$ is the usual L^p-norm. Let $X : \mathbb{W} \to \mathbb{R}^d$ be a smooth Wiener functional in $\bigcap_{k,p} \mathbb{D}^{k,p}$. Let $\Sigma_{ij}^X := \langle DX^i, DX^j\rangle_{\mathbb{H}}$ be the Malliavin covariance matrix. The classical Malliavin calculus studies the problem that under

[1]Corresponding author.

what conditions on X, the law of X has a smooth density with respect to the Lebesgue measure. In particular, as application, Malliavin gave a probabilistic proof for the celebrated Hörmander's hypoellipticity theorem (cf. [16,17]). Nowadays, the Malliavin calculus, as a kind of infinite dimensional analysis, has been extensively used in many fields such as heat kernel estimates, large deviation theory, financial mathematics, numerical calculations, and so on (cf. [6,12,14]).

On the other hand, in the studies of the ergodicity of stochastic dynamical systems, the notion of strong Feller property plays a crucial role (cf. [9]), which relates to the following problem: Let Λ be a metric space and $(X_\lambda)_{\lambda\in\Lambda}$ a random field. We want to seek conditions on X_λ so that for any $f \in \mathcal{B}_b(\mathbb{R}^d)$ (the space of bounded measurable functions),

$$\lambda \mapsto \mathbb{E} f(X_\lambda) \quad \text{is continuous.}$$

In many cases, it is difficult to verify. As we know, if $X_t(x)$ is the solution of an SDE, there are many ways to derive the strong Feller property of $P_t f(x) := \mathbb{E} f(X_t(x))$. For examples, Bismut–Elworthy–Li's formula provides an explicit formula for $\nabla P_t f(x)$ (cf. [11]). Moreover, F. Y. Wang's Hanarck inequality, which gives some quantitive estimate of $P_t f(x)$ for finite and infinite dimensional systems, can also be used to derive the strong Feller property (cf. [22]).

In the framework of the Malliavin calculus, the above problem can be introduced as follows. The celebrated Bouleau–Hirsch's criterion says that if $X_\lambda \in \mathbb{D}^{1,p}$ for some $p > 1$ and the Malliavin covariance matrix $\Sigma^X_\lambda := \Sigma^{X_\lambda}$ is invertible almost surely, then the law of X_λ is absolutely continuous with respect to the Lebesgue measure (cf. [17]). But we have no information about the regularity of the density ρ_λ. In order to obtain such information, one usually needs a strong hypothesis $(\Sigma^X_\lambda)^{-1} \in \bigcap_{p\geq 1} L^p$. If this is true and we work with a diffusion process, then the semigroup of the diffusion has a "regularization effect." The question is: is it possible to emphasize a regularization effect under a weaker hypothesis "$\det(\Sigma^X_\lambda) > 0$ almost surely"? The answer is yes. In fact, Bogachev [3], Corollary 9.6.12, has already shown the following result: Let X_n and X be d-dimensional random variables in $\mathbb{D}^{1,p}$ so that $X_n \to X$ in $\mathbb{D}^{1,p}$. If $p \geq d$ and for almost all ω,

$$\bigl\{D_h X(\omega), h \in \mathbb{H}\bigr\} = \mathbb{R}^d,$$

then the laws of X_n converge to the law of X in total variation. Notice that $\det(\Sigma^X(\omega)) > 0$ implies the above condition, which can be seen as follows: Suppose that $\{D_h X(\omega), h \in \mathbb{H}\} \neq \mathbb{R}^d$, then there is a non-zero vector $v = (v_1, \dots, v_d) \in \mathbb{R}^d$ such that

$$\bigl\langle D_h X(\omega), v\bigr\rangle_{\mathbb{R}^d} = 0, \quad \forall h \in \mathbb{H} \Rightarrow \sum_i v_i D X^i(\omega) = 0 \Rightarrow \Sigma^X(\omega) v = 0 \Rightarrow \det\bigl(\Sigma^X(\omega)\bigr) = 0.$$

This criterion recently was reproven by Malicet and Poly in [15], Corollary 2.2, by using another argument (see also Bally and Caramellino [2], Corollary 2.16). We also mention that the convergence of the densities of random variables has been studied by Ren and Watanabe in [20] under stronger assumptions.

The first aim of this work is to extend Bogachev's result as follows.

Theorem 1.1. *Let $(X_\lambda)_{\lambda\in\Lambda}$ be a family of $\mathbb{R}^d$-valued Wiener functionals over $\mathbb{W}$. Suppose that for some $p > 1$,*

(H1) *$X_\lambda \in \mathbb{D}^{2,p}$ for each $\lambda \in \Lambda$, and $\lambda \mapsto \|X_\lambda\|_{2,p}$ is locally bounded.*
(H2) *$\lambda \mapsto X_\lambda$ is continuous in probability, i.e., for any $\varepsilon > 0$ and $\lambda_0 \in \Lambda$,*

$$\lim_{\lambda\to\lambda_0} \mathbb{P}\bigl(|X_\lambda - X_{\lambda_0}| \geq \varepsilon\bigr) = 0.$$

(H3) *For each $\lambda \in \Lambda$, the Malliavin covariance matrix Σ^X_λ of X_λ is invertible almost surely.*

Then the law of X_λ in $\mathbb{R}^d$ admits a density $\rho_\lambda(x)$ so that $\lambda \mapsto \rho_\lambda$ is continuous in $L^1(\mathbb{R}^d)$.

Remark 1.2. *Our proof is different from [2,3,15] and based on the Sobolev's compact embedding. Compared with [3], our result requires less integrability and continuity assumptions, while more differentiability condition is needed. This can be considered as the case that the differentiability index can compensate the integrability index in infinite dimensional calculus.*

Our another aim of this work is to apply the above criterion to the SDE driven by degenerate jump noises. Let $\mathbb{S}$ be the space of all càdlàg functions from $\mathbb{R}_+$ to $\mathbb{R}^m_+$ with $\ell_0 = 0$ and each component being increasing and pure jump. Suppose that $\mathbb{S}$ is endowed with the Skorohod metric and the probability measure $\mu_{\mathbb{S}}$ so that the coordinate process

$$S_t(\ell) := \ell_t = \big(\ell^1_t, \dots, \ell^m_t\big)$$

is an m-dimensional Lévy process with Laplace transform

$$\mathbb{E}^{\mu_{\mathbb{S}}}\big(\mathrm{e}^{-z \cdot S_t}\big) = \exp\left\{\int_{\mathbb{R}^m_+} \big(\mathrm{e}^{-z\cdot u} - 1\big)\nu_S(\mathrm{d}u)\right\}. \tag{1.1}$$

Consider the following product probability space

$$(\Omega, \mathcal{F}, \mathbb{P}) := \big(\mathbb{W} \times \mathbb{S}, \mathcal{B}(\mathbb{W}) \times \mathcal{B}(\mathbb{S}), \mu_{\mathbb{W}} \times \mu_{\mathbb{S}}\big).$$

If we define W_t and S_t on this probability space, then W_t and S_t are independent, and the subordinate Brownian motion

$$W_{S_t} := \big(W^1_{S^1_t}, \dots, W^m_{S^m_t}\big)$$

is an m-dimensional Lévy process. Below we assume

$$\mathbb{P}\big(\omega \in \Omega : \exists j = 1, \dots, m \text{ and } \exists t > 0 \text{ such that } S^j_t(\omega) = 0\big) = 0, \tag{1.2}$$

which means that S_t is nondegenerate along each direction.

Consider the following SDE driven by W_{S_t}:

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + A\,\mathrm{d}W_{S_t}, \quad X_0 = x, \tag{1.3}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ is a smooth function, $A = (a_{ij})$ is a $d \times m$ constant matrix. Let $H : \mathbb{R}^d \to \mathbb{R}^+$ be a C^∞-function with $\lim_{|x|\to\infty} H(x) = \infty$, which is called a Lyapunov function. We assume that for some Lyapunov function H and $\kappa_1, \kappa_2, \kappa_3 \geq 0$,

$$b(x) \cdot \nabla H(x) \leq \kappa_1 H(x), \tag{1.4}$$

and for all $k = 1, \dots, m$,

$$\left|\sum_i \partial_i H(x) a_{ik}\right|^2 \leq \kappa_2 H(x), \qquad \sum_{ij} \partial_i \partial_j H(x) a_{ik} a_{jk} \leq \kappa_3. \tag{1.5}$$

Under (1.4)–(1.5), X. Zhang in [27], Theorem 3.1, has already proved that SDE (1.3) has a unique solution $X_t(x)$, which defines a Markov process. The associated Markov semigroup is defined by

$$P_t f(x) := \mathbb{E} f\big(X_t(x)\big).$$

We say that (b, A) satisfies a Hörmander's condition at one point $x \in \mathbb{R}^d$ if for some $n = n(x) \in \mathbb{N}$,

$$\mathrm{Rank}\big[A, B_1(x)A, B_2(x)A, \dots, B_n(x)A\big] = d, \tag{1.6}$$

where $B_1(x) := (\nabla b)_{ij}(x) = (\partial_j b^i(x))_{ij}$, and for $n \geq 2$,

$$B_n(x) := (b \cdot \nabla)B_{n-1}(x) - (\nabla b \cdot B_{n-1})(x).$$

Now we can give our main result, which will be proven in Section 3.

Theorem 1.3. *Assume that* (b, A) *satisfies* (1.4)–(1.5) *and Hörmander's condition* (1.6) *at each point* $x \in \mathbb{R}^d$. *Then for any* $t > 0$, *the law of* $X_t(x)$ *is continuous in variable* x *with respect to the total variation distance. In particular, the semigroup* $(P_t)_{t>0}$ *has the strong Feller property, i.e., for any* $t > 0$ *and* $f \in \mathcal{B}_b(\mathbb{R}^d)$,

$$x \mapsto \mathbb{E} f\big(X_t(x)\big) \quad \text{is continuous.}$$

Remark 1.4. *If* $\operatorname{Rank}(A) = d$, *then we can take* $H(x) := |x|^2 + 1$ *so that* (1.4) *becomes*

$$x \cdot b(x) \le \kappa_1\big(|x|^2 + 1\big).$$

In this case, the strong Feller property holds for SDE (1.3) (*cf.* [23,25]).

The topic about the smoothness of the distributional densities of SDEs with jumps has been studied for a long time since the work of Malliavin [16]. We mention the following results:

- By using Girsanov's transformation, Bismut in [5] established an integration by parts formula for Poisson functionals and then used it to study the smoothness of the distributional densities of nondegenerate SDEs with jumps. His idea was systematically developed in the monograph [4].
- In [18], Picard introduced a difference operator argument and derive a new criterion about the smoothness of the distributional densities of Poisson functionals, which is also used to SDEs with jumps. Recently, Ishikawa and Kunita in [13] extended Picard's result to the Wiener–Poisson functional cases. Moreover, Cass [8] studied the SDEs driven by Browian motions and Poisson point processes under Hörmander's conditions. However, the result in [8] does not cover the cases of (1.6) and α-stable noises.
- If $b(x) = Bx$, condition (1.6) is also called Kalman's condition. In this case, Priola and Zabczyk [19] proved the existence of smooth density for the corresponding Ornstein–Uhlenbeck process. In [26], X. Zhang proved the existence of density for SDE (1.3) when b is smooth and Lipschitz continuous. In a special degenerate case, the smoothness of the density is also obtained (cf. [26,27]).

To the best of the authors' knowledge, Theorem 1.3 is the first result about the regularization effect of Lévy noises under *full* Hörmander's conditions. One motivation of our studies comes from the following stochastic oscillators studied in [7,10,21] etc.:

$$\begin{cases} \mathrm{d}z_i(t) = u_i(t)\,\mathrm{d}t, & i = 1, \dots, d, \\ \mathrm{d}u_i(t) = -\partial_{z_i} H(z(t), u(t))\,\mathrm{d}t, & i = 2, \dots, d-1, \\ \mathrm{d}u_i(t) = -[\partial_{z_i} H(z(t), u(t)) + \gamma_i u_i(t)]\,\mathrm{d}t + \sqrt{T_i}\,\mathrm{d}W^i_{S^i_t}, & i = 1, d, \end{cases} \tag{1.7}$$

where $d \ge 3$, $\gamma_1, \gamma_d \in \mathbb{R}$, $T_1, T_d > 0$, and

$$H(z, u) := \sum_{i=1}^{d} \left(\frac{1}{2}|u_i|^2 + V(z_i)\right) + \sum_{i=1}^{d-1} U(z_{i+1} - z_i).$$

The typical examples of V and U are

$$V(z) = \frac{|z|^2}{2}, \qquad U(z) = \frac{|z|^2}{2} + \frac{|z|^4}{4}.$$

The Hamiltonian H describes a chain of particles with nearest-neighbor interaction. We have

Proposition 1.5. *Assume that* $V, U \in C^\infty(\mathbb{R})$ *are nonnegative and* $\lim_{|z|\to\infty} V(z) = \infty$ *so that* H *is a Lyapunov function. If* U *is strictly convex, then* (1.4), (1.5) *and* (1.6) *hold.*

This proposition will be proven in Section 4.

2. Proof of Theorem 1.1

Below, we fix a point $\lambda_0 \in \Lambda$ and a neighbourhood E_{λ_0} of λ_0. We divide the proof into three steps.

(1) Let $\mathrm{GL}(d) \simeq \mathbb{R}^d \otimes \mathbb{R}^d$ be the set of all $(d \times d)$-matrix. Define

$$K_n := \{A \in \mathrm{GL}(d) : \|A\| \le n, \det(A) \ge 1/n\}.$$

Then K_n is a compact subset of $\mathrm{GL}(d)$. Let $\Phi_n \in C^\infty(\mathbb{R}^d \otimes \mathbb{R}^d)$ be a smooth function so that

$$\Phi_n|_{K_n} = 1, \qquad \Phi_n|_{K_{n+1}^c} = 0, \quad 0 \le \Phi_n \le 1.$$

For each $\lambda \in \Lambda$ and $n \in \mathbb{N}$, let us define a finite measure $\mu_{\lambda,n}(\mathrm{d}x)$ by

$$\mu_{\lambda,n}(A) := \mathbb{E}\big[1_A(X_\lambda)\Phi_n\big(\Sigma_\lambda^X\big)\big], \quad A \in \mathcal{B}\big(\mathbb{R}^d\big).$$

Then for each $\varphi \in C_b^\infty(\mathbb{R}^d)$, by [17], Proposition 2.1.4 of p. 100, we have

$$\int_{\mathbb{R}^d} \nabla\varphi(x)\mu_{\lambda,n}(\mathrm{d}x) = \mathbb{E}\big[\nabla\varphi(X_\lambda)\Phi_n\big(\Sigma_\lambda^X\big)\big] = \mathbb{E}\big[\varphi(X_\lambda)\delta\big(\Phi_n\big(\Sigma_\lambda^X\big)\big(\Sigma_\lambda^X\big)^{-1} DX_\lambda\big)\big],$$

where $\nabla = (\partial_1, \ldots, \partial_d)$ and δ is the dual operator of D (also called divergence operator). From this, by (H1) and Hölder's inequality, we derive that

$$\left|\int_{\mathbb{R}^d} \nabla\varphi(x)\mu_{\lambda,n}(\mathrm{d}x)\right| \le \|\varphi\|_\infty C(\lambda, n),$$

where $C(\lambda, n)$ is locally bounded in λ. Hence, $\mu_{\lambda,n}$ is absolutely continuous with respect to the Lebesgue measure (cf. [17]), and in particular, the density $p_{\lambda,n}$ satisfies

$$\int_{\mathbb{R}^d} \big|\nabla p_{\lambda,n}(x)\big|\,\mathrm{d}x \le C(\lambda, n),$$

which implies that $p_{\lambda,n}$ is locally bounded in $\mathbb{W}^{1,1}(\mathbb{R}^d)$ with respect to λ. By Rellich–Kondrachov's compact embedding theorem (cf. [1], Theorem 6.3 of p. 168), $\{p_{\lambda,n}\}_{\lambda \in E_{\lambda_0}}$ is compact in $L^1_{\mathrm{loc}}(\mathbb{R}^d)$, and by Fréchet–Kolmogorov's theorem (cf. [24], Ch. 10), we have

$$\lim_{|y|\to 0} \sup_{\lambda \in E_{\lambda_0}} \int_{B_M} \big|p_{\lambda,n}(x) - p_{\lambda,n}(x+y)\big|\,\mathrm{d}x = 0, \tag{2.1}$$

where $B_M := \{x \in \mathbb{R}^d : |x| \le M\}$ and $M > 0$.

(2) Let $\phi \in C_c^\infty(B_1)$ be a nonnegative smooth function with $\int \phi = 1$. For $\varepsilon > 0$, let

$$\phi_\varepsilon(x) := \varepsilon^{-d}\phi\big(\varepsilon^{-1}x\big).$$

For $f \in \mathcal{B}_b(\mathbb{R}^d)$ with support in B_M, let

$$f_\varepsilon(x) := \int_{\mathbb{R}^d} f(y)\phi_\varepsilon(x-y)\,\mathrm{d}y.$$

Noticing that

$$\begin{aligned}\mathbb{E}\big[\big(f(X_\lambda) - f_\varepsilon(X_\lambda)\big)\Phi_n\big(\Sigma_\lambda^X\big)\big] &= \int_{\mathbb{R}^d} \big(f(y) - f_\varepsilon(y)\big)p_{\lambda,n}(y)\,\mathrm{d}y \\ &= \int_{\mathbb{R}^d} f(y) \int_{\mathbb{R}^d} \big(p_{\lambda,n}(y) - p_{\lambda,n}(y-x)\big)\phi_\varepsilon(x)\,\mathrm{d}x\,\mathrm{d}y,\end{aligned}$$

and in view of $f|_{B_M^c}=0$, we have

$$\begin{aligned}\big|\mathbb{E}\big[\big(f(X_\lambda)-f_\varepsilon(X_\lambda)\big)\Phi_n\big(\Sigma_\lambda^X\big)\big]\big| &\le \|f\|_\infty \int_{B_M}\int_{\mathbb{R}^d}\big|p_{\lambda,n}(y)-p_{\lambda,n}(y-x)\big|\phi_\varepsilon(x)\,\mathrm{d}x\,\mathrm{d}y\\ &\le \|f\|_\infty \sup_{x\in B_\varepsilon}\int_{B_M}\big|p_{\lambda,n}(y)-p_{\lambda,n}(y-x)\big|\,\mathrm{d}y. \end{aligned}\tag{2.2}$$

On the other hand, since $\mathbb{D}^{k,q}=(I-\mathcal{L})^{-k}(L^q)$ for any $q>1$ (Meyer's inequality), where $\mathcal{L}=-\delta D$ is the Ornstein–Uhlenbeck operator, by the interpolation inequality, we have

$$\|DX_\lambda - DX_{\lambda_0}\|_q \le C\|X_\lambda - X_{\lambda_0}\|_q^{1/2}\|X_\lambda - X_{\lambda_0}\|_{2,q}^{1/2},$$

which together with (H1) and (H2) implies that for any $q\in(1,p)$,

$$\lim_{\lambda\to\lambda_0}\|DX_\lambda - DX_{\lambda_0}\|_q = 0.$$

Hence,

$$\lambda\to\Sigma_\lambda^X \quad \text{is continuous in probability.}\tag{2.3}$$

Observe that

$$\begin{aligned}\big|\mathbb{E}\big(f(X_\lambda)-f(X_{\lambda_0})\big)\big| &\le \big|\mathbb{E}\big(f(X_\lambda)-f_\varepsilon(X_\lambda)\big)\big| + \big|\mathbb{E}\big(f(X_{\lambda_0})-f_\varepsilon(X_{\lambda_0})\big)\big| + \mathbb{E}\big|f_\varepsilon(X_\lambda)-f_\varepsilon(X_{\lambda_0})\big|\\ &\le \big|\mathbb{E}\big[\big(f(X_\lambda)-f_\varepsilon(X_\lambda)\big)\Phi_n\big(\Sigma_\lambda^X\big)\big]\big| + 2\|f\|_\infty\mathbb{E}\big|1-\Phi_n\big(\Sigma_\lambda^X\big)\big|\\ &\quad + \big|\mathbb{E}\big[\big(f(X_{\lambda_0})-f_\varepsilon(X_{\lambda_0})\big)\Phi_n\big(\Sigma_{\lambda_0}^X\big)\big]\big| + 2\|f\|_\infty\mathbb{E}\big|1-\Phi_n\big(\Sigma_{\lambda_0}^X\big)\big|\\ &\quad + \|f\|_\infty\int_{B_M}\mathbb{E}\big|\phi_\varepsilon(X_\lambda - y)-\phi_\varepsilon(X_{\lambda_0}-y)\big|\,\mathrm{d}y. \end{aligned}$$

By (2.1), (2.2), (2.3) and taking limits in order $\lambda\to\lambda_0$, $\varepsilon\to 0$ and $n\to\infty$, we obtain

$$\lim_{\lambda\to\lambda_0}\sup_{\|f\|_\infty\le 1, f|_{B_M^c}=0}\big|\mathbb{E}\big(f(X_\lambda)-f(X_{\lambda_0})\big)\big| \le 4\lim_{n\to\infty}\mathbb{P}\big(\Sigma_{\lambda_0}^X\notin K_n\big)\overset{\text{(H3)}}{=}0.\tag{2.4}$$

(3) Lastly, noticing that for any $M>0$,

$$\sup_{\|f\|_\infty\le 1}\big|\mathbb{E}\big(f(X_\lambda)-f(X_{\lambda_0})\big)\big| \le \sup_{\|f\|_\infty\le 1, f|_{B_M^c}=0}\big|\mathbb{E}\big(f(X_\lambda)-f(X_{\lambda_0})\big)\big| + \mathbb{P}\big(|X_\lambda|>M\big)+\mathbb{P}\big(|X_{\lambda_0}|>M\big),$$

by (2.4), Chebyshev's inequality and (H1), we get

$$\lim_{\lambda\to\lambda_0}\sup_{\|f\|_\infty\le 1}\big|\mathbb{E}\big(f(X_\lambda)-f(X_{\lambda_0})\big)\big| = 0.$$

The proof is thus completed by (H1), (H3) and [17], Theorem 2.1.1 of p. 92.

3. Proof of Theorem 1.3

The following lemma is proven in [26], Lemma 2.1.

Lemma 3.1. *For $s>0$, set $\Delta\ell_s^j := \ell_s^j - \ell_{s-}^j$ and*

$$\mathbb{S}_0 := \big\{\ell\in\mathbb{S} : \big\{s : \Delta\ell_s^j > 0\big\} \text{ is dense in } [0,\infty), \forall j=1,\dots,m\big\}.$$

Under (1.2), we have $\mu_{\mathbb{S}}(\mathbb{S}_0)=1$.

Fix $\ell \in \mathbb{S}_0$ and consider the following SDE:

$$\mathrm{d}X_t^\ell(x) = b\big(X_t^\ell(x)\big)\,\mathrm{d}t + A\,\mathrm{d}W_{\ell_t}, \quad X_0^\ell = x. \tag{3.1}$$

The following result is proven in [27], Theorem 3.1.

Theorem 3.2. *Under* (1.4)–(1.5), *there exists a unique solution to SDE* (3.1) *so that for all* $t > 0$,

$$\mathbb{E}\left[\exp\left\{\frac{2\sup_{s\in[0,t]} H(X_s^\ell(x))}{\mathrm{e}^{\kappa_1 t}(\kappa_2|\ell_t|+1)}\right\}\right] \le C_{\kappa_2,\kappa_3}\mathrm{e}^{H(x)}, \tag{3.2}$$

where $C_{\kappa_2,\kappa_3} \ge 1$. *In particular, we have*

$$\mathbb{E}f\big(X_t(x)\big) = \mathbb{E}\big(\mathbb{E}f\big(X_t^\ell(x)\big)|_{\ell=S}\big).$$

For proving the conclusion of Theorem 1.3, by Lemma 3.1, it suffices to show that for each $\ell \in \mathbb{S}_0$ and $t > 0$,

$$\text{the law of } X_t^\ell(x) \text{ is continuous in } x \text{ with respect to the total variation distance.} \tag{3.3}$$

For any $n \in \mathbb{N}$, let $\chi_n(x)$ be a cut-off function on $[0,\infty)$ with

$$\chi_n|_{B_n} = 1, \qquad \chi_n|_{B_{n+1}^c} = 0, \quad 0 \le \chi_n \le 1,$$

and set

$$b_n(x) = b(x)\chi_n\big(H(x)\big).$$

Since $H \in C^\infty(\mathbb{R}^d;\mathbb{R}_+)$ and $\lim_{|x|\to\infty} H(x) = \infty$, we have

$$b_n \in C_b^\infty\big(\mathbb{R}^d\big).$$

Consider the following SDE:

$$\mathrm{d}X_t^n(x) = b_n\big(X_t^n(x)\big)\,\mathrm{d}t + A\,\mathrm{d}W_{\ell_t}, \quad X_0^n = x. \tag{3.4}$$

For fixed $t > 0$ and $n \in \mathbb{N}$, it is easy to see that (H1) and (H2) hold for $x \mapsto X_t^n(x)$. On the other hand, the Malliavin covariance matrix of $X_t^n(x)$ has the following representation (cf. [27], Lemma 4.5):

$$\Sigma_x^{X_t^n} = J_t^n(x)\left(\sum_{k=1}^m \int_0^t K_s^n(x)a_{\cdot k}\big(K_s^n(x)a_{\cdot k}\big)^*\,\mathrm{d}\ell_s^k\right)\big(J_t^n(x)\big)^*,$$

where $J_t^n(x)$ and $K_t^n(x)$ solve the following matrix valued ODEs:

$$J_t^n(x) = I + \int_0^t \nabla b_n\big(X_s^n(x)\big)\cdot J_s^n(x)\,\mathrm{d}s$$

and

$$K_t^n(x) = I - \int_0^t K_s^n(x)\cdot\nabla b_n\big(X_s^n(x)\big)\,\mathrm{d}s.$$

Define

$$B_n^H := \big\{x \in \mathbb{R}^d : H(x) < n\big\}.$$

If (b, A) satisfies Hörmander's condition (1.6) at one point $x \in B_n^H$, then it is easy to see that (b_n, A) also satisfies Hörmander's condition (1.6) at the point $x \in B_n^H$. Thus, from the proof of [26], Theorem 1.1, one sees that $\Sigma_x^{X_t^n}$ is invertible almost surely for $x \in B_n^H$. Using Theorem 1.1, for any $y \in B_n^H$, we have

$$\lim_{x\to y} \sup_{\|f\|_\infty \le 1} \left|\mathbb{E}\left[f\left(X_t^n(x)\right) - f\left(X_t^n(y)\right)\right]\right| = 0. \tag{3.5}$$

Now, for any $x \in B_n^H$, define a stopping time

$$\tau_n(x) := \inf\left\{t \ge 0 : H\left(X_t^\ell(x)\right) \ge n\right\}.$$

By the uniqueness of solutions to SDE, we have

$$X_t^n(x) = X_t^\ell(x), \quad \forall t < \tau_n(x), \text{ a.s.}$$

Let f be a bounded nonnegative measurable function. For any $x, y \in B_n^H$, we have

$$\begin{aligned}
\left|\mathbb{E}\left[f\left(X_t^\ell(x)\right) - f\left(X_t^\ell(y)\right)\right]\right| &\le \left|\mathbb{E}\left[f\left(X_t^\ell(x)\right)1_{t<\tau_n(x)} - f\left(X_t^\ell(y)\right)1_{t<\tau_n(y)}\right]\right| \\
&\quad + \|f\|_\infty \mathbb{P}\left(t \ge \tau_n(x)\right) + \|f\|_\infty \mathbb{P}\left(t \ge \tau_n(y)\right) \\
&= \left|\mathbb{E}\left[f\left(X_t^n(x)\right)1_{t<\tau_n(x)} - f\left(X_t^n(y)\right)1_{t<\tau_n(y)}\right]\right| \\
&\quad + \|f\|_\infty \mathbb{P}\left(t \ge \tau_n(x)\right) + \|f\|_\infty \mathbb{P}\left(t \ge \tau_n(y)\right) \\
&\le \left|\mathbb{E}\left[f\left(X_t^n(x)\right) - f\left(X_t^n(y)\right)\right]\right| \\
&\quad + 2\|f\|_\infty \mathbb{P}\left(t \ge \tau_n(x)\right) + 2\|f\|_\infty \mathbb{P}\left(t \ge \tau_n(y)\right).
\end{aligned}$$

Hence, by (3.5) and (3.2), we obtain

$$\begin{aligned}
\lim_{x\to y} \sup_{\|f\|_\infty \le 1} \left|\mathbb{E}\left[f\left(X_t^\ell(x)\right) - f\left(X_t^\ell(y)\right)\right]\right| &\le 4 \lim_{n\to\infty} \sup_{|x-y|\le 1} \mathbb{P}\left(t \ge \tau_n(x)\right) \\
&\le 4 \lim_{n\to\infty} \sup_{|x-y|\le 1} \mathbb{P}\Big(\sup_{s\in[0,t]} H\left(X_s^\ell(x)\right) \ge n\Big) \\
&\le 4 \lim_{n\to\infty} \frac{1}{n} \sup_{|x-y|\le 1} \mathbb{E}\Big(\sup_{s\in[0,t]} H\left(X_s^\ell(x)\right)\Big) = 0.
\end{aligned}$$

The proof is complete.

4. Proof of Proposition 1.5

Let $x = (z_1, \dots, z_d, u_1, \dots, u_d) \in \mathbb{R}^d \times \mathbb{R}^d$ and define

$$b(x) := b(z, u) := \left(u_1, \dots, u_d, -[\partial_{z_1} H + \gamma_1 u_1], \dots, -\partial_{z_i} H, \dots, -[\partial_{z_d} H + \gamma_d u_d]\right)$$

and

$$A = (a_{i,j}) \quad \text{with } a_{d+1,d+1} = \sqrt{T_1}, \qquad a_{2d,2d} = \sqrt{T_d}, \qquad a_{i,j} = 0 \quad \text{for other } i, j.$$

Clearly,

$$b(x) \cdot \nabla H(x) = -\gamma_1^2 u_1^2 - \gamma_d^2 u_d^2 \le 0.$$

Moreover,

$$\sum_i \partial_i H(x) a_{i,d+1} = \sqrt{T_1} u_1, \qquad \sum_i \partial_i H(x) a_{i,2d} = \sqrt{T_d} u_d$$

and

$$\sum_{ij}\partial_i\partial_j H(x)a_{i,d+1}a_{j,d+1}=T_1,\qquad \sum_{ij}\partial_i\partial_j H(x)a_{i,2d}a_{j,2d}=T_d.$$

Hence, (1.4) and (1.5) hold.

Let us now check (1.6). Let $\mathcal{V}(x)$ be a vector field defined by

$$\begin{aligned}\mathcal{V}(x):=\mathcal{V}(z,u)&:=\sum_{i=1}^{d}b_i(z,u)\partial_{z_i}+\sum_{i=1}^{d}b_{i+d}(z,u)\partial_{u_i}\\&=\sum_{i=1}^{d}u_i\partial_{z_i}-\big(\gamma_1u_1+V'(z_1)-U'(z_2-z_1)\big)\partial_{u_1}\\&\quad-\sum_{i=2}^{d-1}\big(V'(z_i)-U'(z_{i+1}-z_i)+U'(z_i-z_{i-1})\big)\partial_{u_i}\\&\quad-\big(\gamma_du_d+V'(z_d)+U'(z_d-z_{d-1})\big)\partial_{u_d}.\end{aligned}$$

Here the prime denotes the differential. Set $\mathcal{U}_0:=\partial_{u_1}$ and define recursively

$$\mathcal{U}_n:=[\mathcal{U}_{n-1},\mathcal{V}]=\mathcal{U}_{n-1}\mathcal{V}-\mathcal{V}\mathcal{U}_{n-1},\quad n\in\mathbb{N}.$$

By direct calculations, we have

$$\begin{aligned}&\mathcal{U}_1=\partial_{z_1}-\gamma_1\partial_{u_1},\\&\mathcal{U}_2=U''(z_2-z_1)\partial_{u_2}+\big(\gamma_1^2-V''(z_1)-U''(z_2-z_1)\big)\partial_{u_1}-\gamma_1\partial_{z_1}\end{aligned}$$

and

$$\begin{aligned}\mathcal{U}_3=&\,U''(z_2-z_1)\partial_{z_2}+\big(\gamma_1^2-V''(z_1)-U''(z_2-z_1)\big)\partial_{z_1}\\&+\big(\gamma_1V''(z_1)+\gamma_1U''(z_2-z_1)+u_1V^{(3)}(z_1)+(u_2-u_1)U^{(3)}(z_2-z_1)\big)\partial_{u_1}\\&+\big((u_1-u_2)U^{(3)}(z_2-z_1)-\gamma_1U''(z_2-z_1)\big)\partial_{u_2}.\end{aligned}$$

By induction, it is easy to see that for any $k=1,\dots,d-2$,

$$\begin{cases}\mathcal{U}_{2k}=U''(z_{k+1}-z_k)\cdots U''(z_2-z_1)\partial_{u_{k+1}}+\sum_{i=1}^{k}(f_{ki}(x)\partial_{z_i}+g_{ki}(x)\partial_{u_i}),\\ \mathcal{U}_{2k+1}=U''(z_{k+1}-z_k)\cdots U''(z_2-z_1)\partial_{z_{k+1}}+\sum_{i=1}^{k}(\tilde f_{ki}(x)\partial_{z_i}+\tilde g_{ki}(x)\partial_{u_i})+h_k(x)\partial_{u_{k+1}},\end{cases}$$

where $f_{ki},g_{ki},\tilde f_{ki},\tilde g_{ki},h_k$ are smooth functions. Since $U''>0$, we have

$$\partial_{u_1},\partial_{z_1},\dots,\partial_{u_{d-1}},\partial_{z_{d-1}}\in\operatorname{span}\{\mathcal{U}_0,\mathcal{U}_1,\dots,\mathcal{U}_{2d-3}\}.\tag{4.1}$$

On the other hand, since

$$[\partial_{u_d},\mathcal{V}]=\partial_{z_d}-\gamma_d\partial_{u_d},$$

by (4.1) we further have

$$\partial_{u_1},\partial_{z_1},\dots,\partial_{u_d},\partial_{z_d}\in\operatorname{Span}\big\{\mathcal{U}_0,\mathcal{U}_1,\dots,\mathcal{U}_{2d-3},\partial_{u_d},[\partial_{u_d},\mathcal{V}]\big\},$$

which means that (1.6) holds.

Acknowledgements

The authors deeply thank the referees for their very useful suggestions so that Theorem 1.1 is improved. Zhao Dong and Xuhui Peng are supported by 973 Program, No. 2011CB808000 and Key Laboratory of Random Complex Structures and Data Science, No. 2008DP173182, NSFC, Nos.: 10721101, 11271356, 11371041. Xicheng Zhang is supported by NNSFs of China (Nos. 11271294, 11325105).

References

[1] R. A. Adams and J. F. Fournier. *Sobolev Spaces*, 2nd edition. Academic Press, Amsterdam, 2003. MR2424078
[2] V. Bally and L. Caramellino. On the distances between probability density functions. *Electron. J. Probab.* **19** (2014) 1–33.
[3] V. I. Bogachev. *Differentiable Measures and the Malliavin Calculus*. *Mathematical Surveys and Monographs* **164**. Amer. Math. Soc., Providence, RI, 2010. MR2663405
[4] K. Bichteler, J. B. Gravereaux and J. Jacod. *Malliavin Calculus for Processes with Jumps*. Gordan and Breach Science Publishers, New York, 1987. MR1008471
[5] J. M. Bismut. Calcul des variations stochastiques et processus de sauts. *Z. Wahrsch. Verw. Gebiete* **63** (1983) 147–235. MR0701527
[6] J. M. Bismut. *Large Deviations and the Malliavin Calculus*. Birkhäuser, Boston, 1984. MR0755001
[7] P. Carmona. Existence and uniqueness of an invariant measure for a chain of oscillators in contact with two heat baths. *Stochastic Process. Appl.* **117** (8) (2007) 1076–1092. MR2340880
[8] T. Cass. Smooth densities for stochastic differential equations with jumps. *Stochastic Process. Appl.* **119** (5) (2009) 1416–1435. MR2513114
[9] G. Da Prato and J. Zabczyk. *Ergodicity for Infinite Dimensional Systems*. *London Math. Society Lecture Notes Series* **229**. Cambridge Univ. Press, Cambridge, 1996. MR1417491
[10] J. P. Eckmann and M. Hairer. Non-equilibrium statistical mechanics of strongly anharmonic chains of oscillators. *Comm. Math. Phys.* **212** (1) (2000) 105–164. MR1764365
[11] K. D. Elworthy and X. M. Li. Formulae for the derivatives of heat semigroups. *J. Funct. Anal.* **125** (1994) 252–286. MR1297021
[12] E. Fournié, J. M. Lasry, J. Lebuchoux, P. L. Lions and N. Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance Stoch.* **3** (1999) 391–412. MR1842285
[13] Y. Ishikawa and H. Kunita. Malliavin calculus on the Wiener–Poisson space and its application to canonical SDE with jumps. *Stochastic Process. Appl.* **116** (2006) 1743–1769. MR2307057
[14] S. Kusuoka and D. Stroock. Applications of the Malliavin calculus. I. In *Stochastic Analysis (Katata/Kyoto, 1982)* 271–306. *North-Holland Math. Library* **32**. North-Holland, Amsterdam, 1984. MR0780762
[15] D. Malicet and G. Poly. Properties of convergence in Dirichlet structures. *J. Funct. Anal.* **264** (2013) 2077–2096. MR3029147
[16] P. Malliavin. Stochastic calculus of variations and hypoelliptic operators. In *Proceedings of the International Symposium on Stochastic Differential Equations (Res. Inst. Math. Sci., Kyoto Univ., Kyoto, 1976)* 195–263. Wiley, New York–Chichester–Brisbane, 1978. MR0536013
[17] D. Nualart. *The Malliavin Calculus and Related Topics*. Springer-Verlag, New York, 2006. MR2200233
[18] J. Picard. On the existence of smooth densities for jump processes. *Probab. Theory Related Fields* **105** (1996) 481–511. MR1402654
[19] E. Priola and J. Zabczyk. Densities for Ornstein–Uhlenbeck processes with jumps. *Bull. Lond. Math. Soc.* **41** (2009) 41–50. MR2481987
[20] J. Ren and S. Watanabe. A convergence theorem for probability densities and conditional expectations of Wiener functionals. In *Dirichlet Forms and Stochastic Processes* 335–344. de Gruyter, Berlin, 1995. MR1366448
[21] L. Rey-Bellet and L. E. Thomas. Asymptotic behavior of thermal nonequillibrium steady states for a driven chain of anharmonic oscillators. *Comm. Math. Phys.* **215** (1) (2000) 1–24. MR1799873
[22] F. Y. Wang. *Harnack Inequalities for Stochastic Partial Differential Equations*. Springer, New York, 2013. MR3099948
[23] F. Y. Wang and J. Wang. Harnack inequalities for stochastic equations driven by Lévy noise. *J. Math. Anal. Appl.* **410** (1) (2014) 513–523. MR3109860
[24] K. Yosida. *Functional Analysis*. Springer-Verlag, Berlin, 1979.
[25] X. Zhang. Derivative formula and gradient estimate for SDEs driven by α-stable processes. *Stochastic Process. Appl.* **123** (4) (2013) 1213–1228. MR3016221
[26] X. Zhang. Densities for SDEs driven by degenerate α-stable processes. *Ann. Probab.* **42** (5) (2014) 1885–1910. MR3262494
[27] X. Zhang. Fundamental solution of kinetic Fokker–Planck operator with anisotropic nonlocal dissipativity. *SIAM J. Math. Anal.* **46** (3) (2014) 2254–2280. MR3225504